



Une archive ouverte des publications en InfoCom

Gabriel Gallezot, Ghislaine Chartron, Jean-Max Noyer

► To cite this version:

Gabriel Gallezot, Ghislaine Chartron, Jean-Max Noyer. Une archive ouverte des publications en InfoCom. Place et enjeux des revues pour la recherche en infoCom (SFSIC), Mar 2002, Nice, France. sic_00000025v2

HAL Id: sic_00000025

https://archivesic.ccsd.cnrs.fr/sic_00000025v2

Submitted on 17 May 2002

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Titre : Une archive ouverte des publications en InfoCom

Auteurs :

Gallezot Gabriel

*GRESI/Enssib,
17-21, boulevard du 11 Novembre 1918
69 623 Villeurbanne*

gabriel.gallezot@free.fr

Chartron Ghislaine

*Urfist-Paris et GRESI/Enssib.
URFIST-Paris
17 rue des Bernardins
75005 Paris.*

chartron@cnam.fr

Noyer Jean-Max

*Université Paris 7, CCI / CRECI,
2, place Jussieu, 75005 Paris.
École militaire de Saint-Cyr Coëtquidan*

jean-max.noyer@uhb.fr

Résumé :

Dans une démarche constructive, les auteurs proposent la création d'une archive ouverte en InfoCom. Après un rappel sur l'origine, les mouvements et les possibilités offertes par les archives ouvertes (Open Archive), nous voulons dresser avec l'ensemble de la communauté InfoCom les contours d'un projet intellectuel favorisant le développement scientifique de notre communauté.

Titre : Une archive ouverte des publications en InfoCom

Auteurs :

Résumé :

Dans une démarche constructive, les auteurs proposent la création d'une archive ouverte en InfoCom. Après un rappel sur l'origine, les mouvements et les possibilités offertes par les archives ouvertes (Open Archive), nous voulons dresser avec l'ensemble de la communauté InfoCom les contours d'un projet intellectuel favorisant le développement scientifique de notre communauté.

Problématisation:

Les modes de publications des résultats de la recherche ou rendant compte de l'état d'avancement de la recherche sont multiples. Articles, rédaction de projets, monographies, communications à des colloques et séminaires (ces derniers ayant eux-mêmes des statuts et fonctions différents)...et ce, sous forme papier et à présent sous forme numérique. Ces objets éditoriaux sont de forme et de nature variables, leur mode de légitimation et de validation très différents. Leur fonction dans le processus de production des savoirs, leurs usages collectifs ou singuliers sont aussi multiples, suivant les communautés, leur taille, leur épistémologie, leurs us et coutumes. L'arrivée du numérique et le développement des nouvelles mémoires entraînent aussi un certain nombre de transformations et soulèvent à la fois des difficultés et certaines opportunités.

Dans le débat qui nous préoccupe, il nous semble déceler quatre questions fondamentales.

La première concerne la "représentation" des revues constitutives de notre discipline. Quels sont les critères qui permettent leur identification, leur évaluation, leur positionnement réelle ou supposée des revues dans l'espace institutionnel, leur charge symbolique...? La discussion concernant ces critères, la nécessité de parvenir à une connaissance plus fine du travail des revues entre elles nous semblent particulièrement importants, d'autant plus que l'hétérogénéité de notre champ est grande. Rendre donc visible selon des niveaux d'échelles à déterminer les organisations internes des dynamiques, des échanges, des références est donc souhaitable. Aussi, la réflexion proposée par les organisateurs du colloque est légitime et nécessaire.

Cela nous conduit à la deuxième question introduisant la contribution que nous voudrions apporter à ce débat. Est-il possible, de créer les conditions permettant de faire émerger, à partir des revues existantes et éventuellement plus en amont, à partir des propositions d'auteurs (sous forme par exemple de pré-publications) une représentation plus fidèle de l'activité collective et donc d'améliorer la mise en visibilité et peut-être de susciter une évaluation plus collective, des dynamiques entre auteurs, entre thématiques, entre laboratoires?

Dans la perspective numérique, existe-t-il de nouveaux modes éditoriaux qui pourraient contribuer à un tel objectif, à rendre visibles jusqu'à un certain point, les dynamiques amont et aval d'un texte qui contextualisent (ne serait-ce que pour partie) le travail et la vie des articles ?¹ Il s'agit donc de développer des visibilités (en partie mesurables) plus vastes, rendant compte des processus amont et aval des productions de savoirs à travers les articles (post-publications et pré-publications), du nombre approché de lecteurs, des réseaux de citations, de co-citations...

Ces nouveaux modes éditoriaux, quels sont-ils ? C'est là notre troisième question.

Ils sont essentiellement recouverts par le vocable "archives ouvertes" et ont tous en commun de permettre un archivage numérique offrant des modes d'exploitation de ces fonds documentaires en vue de nouvelles pratiques et notamment de nouvelles pratiques cartographiques (conceptuelles, d'auteurs, (auteur cité, auteur citant) de revues (revue citée / revue citante ...)

¹ Nous renvoyons ici au travail sur la "Citabilité" en général de Jacques Derrida à Leysderdoff en passant par Solla Price, Garfield, De Landa, Latour, Courtial, Baudouin Jurdant, Lecoadic, Noyer, Alliez, et bien d'autres encore.

Notre quatrième question interroge les diverses manières dont nous pouvons mettre en place et utiliser ces dispositifs porteurs d'une grande valeur ajoutée éditoriale, c'est à dire rendant visibles les structurations dynamiques du champ, améliorant le système hérité du "peer-review" (en lui donnant des moyens accrus de fonder sa pertinence, tout en ne pénalisant pas les options minoritaires, marginales ou émergentes).

C'est donc à partir de cette réflexion générale et sous les contours d'une " archive ouverte " construite avec les revues que nous vous proposons de dresser ensemble un projet intellectuel favorisant le développement de notre communauté.

I – Le mouvement des archives ouvertes dans la communication scientifique: ses origines

La création des archives ouvertes est née en réaction au paradoxe de l'édition scientifique : alors que le chercheur offre ces résultats, la communauté scientifique trouve un certain nombre d'obstacles pour lire ces publications, revues non présentes dans les bibliothèques de proximité, sélection restrictive effectuée par les bibliothèques contraintes par leurs budgets, dispersion des revues... Pourtant la majorité des chercheurs accèdent aujourd'hui à Internet et pourraient se saisir de cette opportunité pour partager plus facilement les résultats de leurs travaux. De plus, l'inflation des écrits et la nécessité d'une validation dans un temps plus court pour certains résultats (souvent expérimentaux) se heurtent aux limites du modèle traditionnel de publication.

Aussi, n'existe-t-il pas une solution qui permettrait de faire circuler de façon plus ouverte la littérature scientifique, un moyen d'accéder à un corpus exhaustif de textes évalués, téléchargeable et disponible à tout moment, corpus où les articles pourraient être reliés et recherchés quelque soit le comité d'évaluation ayant travaillé en amont ? Ce modèle n'est plus une utopie, il est déjà à l'œuvre dans certaines communautés, supporté par le mouvement des "archives ouvertes" réunissant des chercheurs de différentes disciplines et des informaticiens désireux d'expérimenter ces nouveaux modes de circulation des savoirs (voir en particulier S. Harnad).

II– Qu'est qu'une archive ouverte ?

Une " archive ouverte " est la traduction littérale d'Open Archive. Le terme " archive " est communément employé pour désigner la préservation de documents sur une longue durée et par extension la conservation de documents anciens, ici il désigne des archives vivantes, un entrepôt d'informations. De même le terme " ouverte ", bien qu'il renvoie aussi à la notion de consultation libre et gratuite, a pour sens dans ce contexte l'ouverture de l'architecture technique, des protocoles communs qui facilitent l'accès au contenu. Si l'acception française de l'expression " archive ouverte " n'est pas exempte d'ambiguïtés, son signifiant repose sur des réalisations précises qui permettent de mieux l'appréhender (cf. Part III). Une archive ouverte peut se définir simplement comme un réservoir de publications électroniques. Le terme est fortement associé au projet " Open Archives Initiative "² et à la communauté " E-

² <http://www.openarchives.org>

prints”³ deux mouvements importants ces dernières années dans domaine de la communication scientifique électronique. Le premier se préoccupe de l’interopérabilité des archives en établissant des protocoles liés aux métadonnées (basé sur le Dublin Core) et au processus de recherche d’information. Le second encourage et favorise l’auto-archivage des publications savantes par les auteurs/organismes notamment à travers le logiciel gratuit eprints.org⁴. Les principaux enjeux associés aux archives ouvertes sont : un accès ouvert et simplifié aux publications scientifiques, une recherche d’information décuplée (interopérabilité possible entre les archives) et une mémoire scientifique commune.

III – Les différents types d’archives ouvertes

Deux types d’archives se sont développés ces dernières années: les archives dites de pré-publications et celles de post-publications. Les premières sont des bases d’articles dans leur première version et se situent en amont du processus d’évaluation par les pairs, les secondes sont des bases d’articles déjà publiés dans les revues et se situent donc en aval de l’évaluation par les pairs.

Dans les faits, la montée en charge de chacun de ces deux orientations entraîne souvent leur convergence. A titre d’exemple, on peut citer l’une des bases les plus connues et les plus importantes, à savoir la base ArXiv de P. Ginsparg⁵, une archive ouverte en physique (théorique au départ, puis extension à d’autres spécialisations de la physique) créée en 1991. Mais le modèle initié par la physique a connu aussi des extensions non négligeables dans d’autres disciplines: psychologie et neuroscience (archive Cogprints⁶ de S. Harnad, créée en 1997), économie (WoPec⁷, archive ouverte de "working papers" en économie, créée en 1996), différentes archives également plus circonscrites aux universités, à un réseau de chercheurs rendant ainsi plus visibles les travaux publiés.

Certains expérimentent également des fonctionnalités avancées, liées à cette nouvelle forme de communication. Par exemple le "commentaire ouvert par les pairs" dont le principe est d’ouvrir l’article aux commentaires de toute une communauté par courrier électronique, il s’agit de renforcer la controverse ; une autre fonctionnalité avancée concerne l’analyse des citations des articles archivés (mise en relation des différents articles cités dans les bibliographies des textes, quantification des consultations...)

Ainsi, au regard des différentes réalisations d’archives et de leurs projets connexes, la notion d’archive ouverte est suffisamment polymorphe et versatile pour impulser une configuration adaptée à chaque communauté scientifique.

IV – Comment cela fonctionne-t-il ?

³ <http://www.eprints.org/>

⁴ Le logiciel d’auto-archivage, **eprints.org**, a été développé au Département d’Electronique et d’Informatique de l’Université de Southampton.

⁵ <http://xxx.lanl.gov> et <http://fr.arXiv.org> , En 11 ans plus de 185 000 textes ont été déposés, http://xxx.lanl.gov/cgi-bin/show_monthly_submissions [consulté le 14 janvier 2002]

⁶ <http://cogprints.soton.ac.uk>, en 5 ans plus de 4000 textes ont été déposés [consulté le 14 janvier 2002]

⁷ <http://netec.mcc.ac.uk/WoPEc.html>, créé en 1996, environ 50 000 documents, articles de revue + Working papers. Tous les articles sont téléchargeables mais pas nécessairement en libre accès.

ArXiv, première archive créée par une communauté scientifique, fournit un exemple à grande échelle:

- le chercheur envoie son texte à l'archive par courrier électronique ou interface web ;
- un (ou des) modérateur scientifique vérifie la recevabilité du texte qui n'est pas une validation scientifique de haut niveau (toujours assumée par les revues), mais un premier filtrage écartant tout texte posant problème.
- les autres chercheurs peuvent rechercher l'article à partir d'une interface d'interrogation classique sur le web (formulaires) ou être informé par courrier électronique de sa présence dans l'archive (principe d'alerte). Ils le téléchargent sur leur ordinateur pour le lire (impression possible) ;
- la communauté scientifique peut commenter l'article, et l'auteur peut réviser son article et remplacer la 1^{er} version. Ce processus peut être reproduit plusieurs fois ;
- l'article peut être soumis à une revue. Certaines revues permettent une soumission sous la forme d'un numéro de document de l'archive. Les évaluateurs vont sur l'archive pour lire l'article ;
- L'article est alors accepté ou rejeté par la revue. S'il est rejeté, il peut être soumis à une autre revue (après révision ou non), s'il est accepté, l'auteur le corrige éventuellement selon les commentaires des évaluateurs ;
- L'article est publié dans la revue et dans la plupart des cas déposé dans l'archive.

Dans ce scénario lié aux sciences de la nature, l'archive assure un rôle de circulation anticipé des écrits avec des possibilités d'améliorations cumulatives des articles. Mais d'autres scénarios peuvent être envisagés avec les éléments suivants :

- L'archive peut être aussi constituée de textes déjà évalués et publiés dans des revues depuis longtemps (accord avec les revues, négociation de l'auteur).
- Il peut exister un temps de latence entre la publication dans une revue et le dépôt dans l'archive, permettant notamment à la revue de ne pas déstabiliser son modèle économique.
- Plutôt que de déposer son article sur une archive centralisée, le logiciel Eprint.org permet le dépôt sur l'archive de son choix (de son institution, de sa communauté...) en proposant l'interopérabilité de celles-ci.

V – La question du droit d'auteur

La question fréquemment posée à propos du dépôt dans les archives relève du droit d'auteur : ai-je le droit de publier mon article dans une revue et dans une archive?

Harnad et Oppenheim⁸ proposent des solutions pour régler cette question. La première est de ne pas donner l'exclusivité de l'articles aux revues⁹, la deuxième est d'auto-archiver sur le réseau une version avant de le soumettre aux éditeurs, laisser les experts évaluer, et ajouter un appendice correctif au texte sur le réseau. La troisième plus consensuelle consiste à négocier avec les éditeurs des revues, qui peuvent fournir directement à l'archive les articles publiés après un temps de latence restant à définir.

⁸ “ Lecture et écriture scientifique “dans le ciel” : Une anomalie post - gutenbergiennne et comment la résoudre ” [6. Comment contourner légalement un copyright restrictif (la “ stratégie d'Harnad/Oppenheim ” Re : Should Publishers Offer Free-Access Services? <http://www.cogsci.soton.ac.uk/~harnad/Hypermail/Amsci/0542.html>], novembre 2001, http://www.text-e.org/conf/index.cfm?ConfText_ID=7 ..

⁹ <http://cogprints.soton.ac.uk/copyright.html>.

VI – Une archive ouverte en InfoCom ?

L'expérimentation d'une archive ouverte d'articles scientifiques en sciences de l'information et de la communication devrait mettre à disposition de tous les chercheurs en SIC un grand nombre de documents diversifiés et favoriser les échanges pluridisciplinaires. Ce corpus illustrerait ainsi la richesse de nos approches respectives et de nos terrains. A la place d'une conformation uniforme des savoirs liés à l'appauvrissement du nombre de revues consultables par chacun, l'archive ouverte propose une cartographie des connaissances, des thématiques scientifiques qui animent nos recherches. Par l'identification des problématiques majeures traduites par les publics, elle concourrait à la structuration de notre communauté en laissant des traces patrimoniales pour une histoire de notre discipline.

Plus prosaïquement, cette archive ouverte faciliterait l'accès et la recherche de savoirs, elle permettrait une plus grande visibilité des revues où l'on publie (post-publication), elle inscrirait la publication des résultats de la recherche dans un temps plus court également. Elle permettrait aussi d'expérimenter les apports éventuels de la technique dans ce nouveau modèle de publication scientifique notamment en terme de création de connaissances.

La faisabilité d'un tel projet est liée à la disponibilité d'un service d'édition électronique compétent et stable. Le site de l'archive doit bénéficier d'une logistique technique performante (réseau, suivi technologique, sécurité des accès...). Les premières réflexions menées conduisent à privilégier un partenariat avec le service ouvert récemment par le CNRS, le CCSD (Centre pour la Communication Scientifique Directe¹⁰) dont la vocation ne se limite pas à la physique ou autres sciences de la nature, mais s'ouvre au contraire aux projets des sciences humaines et sociales.

L'expérimentation pourrait ainsi facilement démarrer, tout en envisageant progressivement d'autres lieux éventuels d'hébergement souhaités par la communauté. C'est dans cet esprit, lors de ce colloque, que nous souhaitons recueillir l'intérêt de la communauté SIC sur la construction et l'alimentation d'une telle archive ouverte.

¹⁰ <http://ccsd.cnrs.fr/>

Références

- Alliez E., *La signature du monde*, Édition du Cerf, Paris, 1993
- Bosc H. "La communication scientifique revue et corrigée par internet." 2001, <http://www.tours.inra.fr/tours/docs/comsci.htm>
- Callon M., Law J., Rip A., « Mapping the dynamics of science and technology », London, MacMillan, 1986
- Chartron G. "Les archives ouvertes dans la communication scientifiques." 2002, <http://www.ccr.jussieu.fr/urfist/archives-ouvertes.html>.
- Chartron G. (dir) *Les mutations de l'édition de recherche, Internet et les chercheurs*, Cercle de la Librairie (à paraître printemps 2002).
- Chartron G., Salaün J.M., La reconstruction de l'économie politique des publications scientifiques, *Bulletin des Bibliothèques de France*, n°2, t.45, 2000, pp. 32-42
- Courtial J.P., Callon M., Penan H., *Introduction à la scientométrie : de la bibliométrie à la veille technologique*, Paris, *Anthropos- economica* 1990.
- Crawford S.Y., Hurd J.M., Weller A.C., From print to Electronic, the transformation of Scientific communication, ASIS, 1996
- De la Vega J.F., La communication scientifique à l'épreuve de l'Internet, l'émergence d'un nouveau modèle, Presses de l'Enssib, 2000.
- De Landa M., *A Thousand years of non-linear history*, 1997 Ed. Swreve
- Derrida J., *Limited and co*, Édition Galilée, Paris, 1990
- Garfield E. voir le site " Cybermetrics " : <http://www.cindoc.csic.es/cybermetrics/>
- Ginsparg P., "First Steps Towards Electronic Research Communication", *Computers in Physics*, vol.8, n°4, 1994, repris dans *Solaris* n°3, <http://www.info.unicaen.fr/bnum/jelec/Solaris/d03/3ginspar.html>
- Harnad S. "Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry." *Psychological Science* 1, 1990: 342-343.
- Harnad S., "Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge", *Public Access Computer Systems Review* 2 (1), 1991, pp.39-53.
- Hitchcock S., Carr L., Jiao Z., Bergmark D., Hall W., Lagoze C., Harnad S., "Developping services for open eprint archives: globalisation, integration and the impact of links", 5th ACM Conference on Digital Libraries, juin 2000, version pré-publication, <http://www.cogsci.soton.ac.uk/~harnad/Papers/Harnad/harnad00.acm.htm>
- Krichel T. "WoPEc ; Electronic working papers in Economics Services." *Ariadne*, 1997, <http://www.ariadne.ac.uk/issue8/wopec/intro.html>
- Latour B., Teil G., « The hume machine : can associative networks do more than formal rules ? », *Stanford Humanities Review*, 4(2). P47-66
- Le Coadic Y., *La science de l'information*, Paris, PUF, Coll. Que sais-je, 1994.
- Les publications d'Harnad à l'adresse <http://cogprints.ecs.soton.ac.uk/~harnad/>
- Mutschke P., Quan Haase A., « Collaboration and cognitive structures in social science research fields » , in *Scientometrics*, Vol. 52, N°3, 2001
- Nelson M. L. "Smart Objects and Open Archives." *D-Lib Magazine* 7.2 , 2001, <http://www.dlib.org/dlib/february01/nelson/02nelson.html>
- Noyer J.M., « Pourquoi la scientométrie, l'infométrie, nous intéressent-elles ? », in *Solaris* N°2 <http://www.info.unicaen.fr/bnum/jelec/Solaris>
- Noyer J.M. " L'économie du savoir ", *Solaris* n°1, 1994, <http://www.info.unicaen.fr/bnum/jelec/Solaris/d01>
- Odlyzko A. M., "Competition and cooperation: Libraries and publishers in the transition to electronic scholarly journals", *Journal of Electronic Publishing* 4(4), June 1999, <http://www.press.umich.edu/jep/>
- Odlyzko, A.M. "The economics of electronic journals". In: Ekman R. and Quandt, R. (Eds) *Technology and Scholarly Communication*. Univ. Calif. Press, 1998. <http://www.research.att.com/~amo/doc/complete.html>
- Oppenheim, C., 2001, The legal and regulatory environment for electronic information. *Infonortics*. <http://www.infonortics.com/publications/legal4.html>
- Pinfield S. "How do physicists Use an E-print Archive?." *D-Lib Magazine* 7.12, 2001, <http://www.dlib.org/dlib/december01/pinfield/12pinfield.html>
- Solla Price D. *Science and Suprascience*, Édition Fayard, 1972 et *Solaris* N°2 <http://www.info.unicaen.fr/bnum/jelec/Solaris>
- Summer T., Buckingham Shum S. "Open Peer Review & Argumentation : Loosening the paper Chains on Journals." *Ariadne* issue 5, 2001, <http://www.ariadne.ac.uk/issue5/jime/intro.html>
- Van de Sompel H., Lagoze C. "The Santa Fe Convention of the Open Archives Initiative." *D-Lib Magazine*, 2000, <http://www.dlib.org/dlib/february00/vandesompel-oai/vandesompel-oai.html>

